**Business Intelligence Final Project: Crimes and Firearms**

**Analysis of Data**

The main dataset we choose is the ‘crime’ dataset from the Crime Data Explorer. It is a website that allows law enforcement and the general public to more easily access Uniform Crime Reporting (UCR) data. Over 18,000 city, university and college, county, state, tribal, and federal law enforcement agencies voluntarily report crime data to the program, and the FBI publishes it in annual reports (1). It has different kinds of reported crimes (violent crime, rape, robbery, aggravated assault, burglary, larceny and few more) from year 1995 to 2017 for all the states, including aggregated number of crimes each year. This dataset provides us the insight of trend of crime depending on the state and recorded years. Also, using this as primary dataset, helps to research, if these types of crimes have dependencies on anything. Like, education, unemployment, number of firearms, age, sex, race, environment or many more factors. As example, we tried to find the correlation between the no of firearms on violent crimes, or homicide and effect of unemployment rate on burglary. By studying these relations and deploy on the real environment can affect the minimization of different levels of crime. Not only this will have effect on our social life, it will certainly have effect on our economic life, overall on the country’s economy.

Keeping this in mind, here comes the next two datasets which we worked on. First one is ‘Firearms’ dataset. It’s fetched from the ATF (alcohol, tobacco firearms and explosives) site. ATF is a law enforcement agency in the United States’ Department of Justice that protects our communities from violent criminals, criminal organizations, the illegal use and trafficking of firearms, the illegal use and storage of explosives, acts of arson and bombings, acts of terrorism, and the illegal diversion of alcohol and tobacco products. We partner with communities, industries, law enforcement, and public safety agencies to safeguard the public we serve through information sharing, training, research, and use of technology. ATF is responsible for enforcing the Gun Control Act (GCA) of 1968, as amended. The GCA regulates the manufacture, importation, distribution, and sale of firearms, and it contains criminal provisions related to the illegal possession, use, or sale of firearms (2). The dataset has a list of different types of registered and sold firearms in all states in a year, aggregated with total number of firearms sold in a state. By researching this, we can have ideas about the firearms business in each state. Also, we get the knowledge of, how it’s effecting the rate suicide by firearms. It shows indirect business application, we can find this after merging with ‘crime’ and ‘unemployment’ data. if we stop the legalization of firearms, it will affect the economy. There would likely be a modest net economic gain if guns disappeared. Gun death and injury-related expenses add up to direct losses of around $10.7 billion (£7 billion) per year, and more than [$200 billion](https://www.motherjones.com/politics/2015/04/true-cost-of-gun-violence-in-america/) (£140 billion) when other factors are taken into account(3). Moreover, it shows a negative effect with the unemployment rate.

‘Unemployment’ is our 3rd dataset, which we downloaded from ‘Bureau of Labor Statistics’ site. It gives a direct idea about the condition of employment in each state and the comparison with different years by parallelly checking different years of data together. We can research it’s effect by merging it with ‘crime’ and ‘firearms’ dataset as discussed before.

**Data Cleaning**

The quality of all the three datasets we used are good. We had some missing values from the ‘crime’ dataset on the variable names ‘rape\_revised’ and missing values in some ‘Year’ and ‘State’ columns as well. ‘rape\_revised’ means penetration, no matter how slight, of the vagina or anus with any body part or object, or oral penetration by a sex organ of another person, without the consent of the victim. Attempts or assaults to commit rape are also included; however, statutory rape and incest are excluded. In December 2011, the UCR program changed its definition of SRS rape to this revised definition. This change can be seen in the UCR data starting in 2013. Any data reported under the older definition of rape will be called "legacy rape". We removed that column as we had another column named ‘rape\_legacy’. And we exactly didn’t use any rape data column for our calculations and decisions. So, we found it better to remove it from the merged dataset. Also, there was no way to determine the missing ‘State’ values in dataset hence we removed it as well.

**Data Merging**

Data merging is the process of combining two or more data sets into a single data set. Most often, this process is necessary when you have raw data stored in multiple files, worksheets, or data tables, that you want to analyze all in one go.

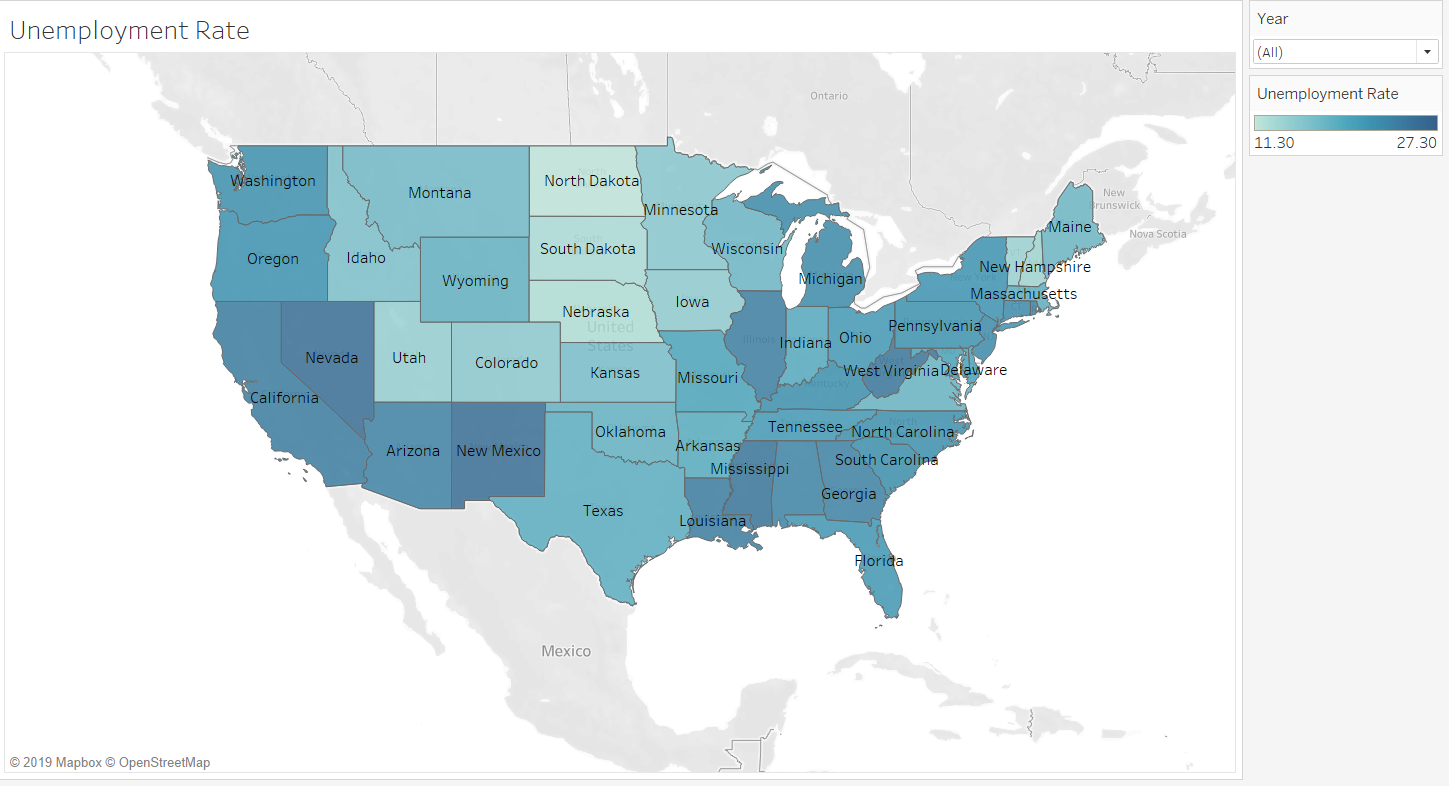
We used 4 different data sets, which are: Unemployment, Crime, Firearms and Mortality by firearms. All the Datasets were cleaned to have information over the period of 4 years from 2014-2017. Year and States are the common elements between each dataset. After cleaning the datasets, we had 4 rows for each state as we wanted to include 4 years of data. We used inner join to merge all the data sets using year and state columns as key columns. There was no data redundancy found and the expected number of rows were created. A total of 26 columns and 203 rows in the data. We realized that the variables such as population of all the states from crime dataset, ‘Total’ firearms registered in each state and ‘Deaths’ from mortality by firearms dataset to compare and contrast. We can also use variables such as ‘rate’ from unemployment dataset and various variables describing crimes in crime dataset to get the effect of increasing or decreasing unemployment rate on Criminal activities in various states.

**Analysis of Visualizations**

The visualizations for this project were done using Tableau 2019. The cleaned dataset was imported to the desktop version of the software and the relevant dimension and metrics were used to create maps and trend lines. The maps below show the rates of unemployment, Violent crimes, Death by firearms and Total firearms. These rates were calculated per capita. Calculated fields were created in Tableau where the respective metric was divided by the population, thus preventing the population bias of the states.

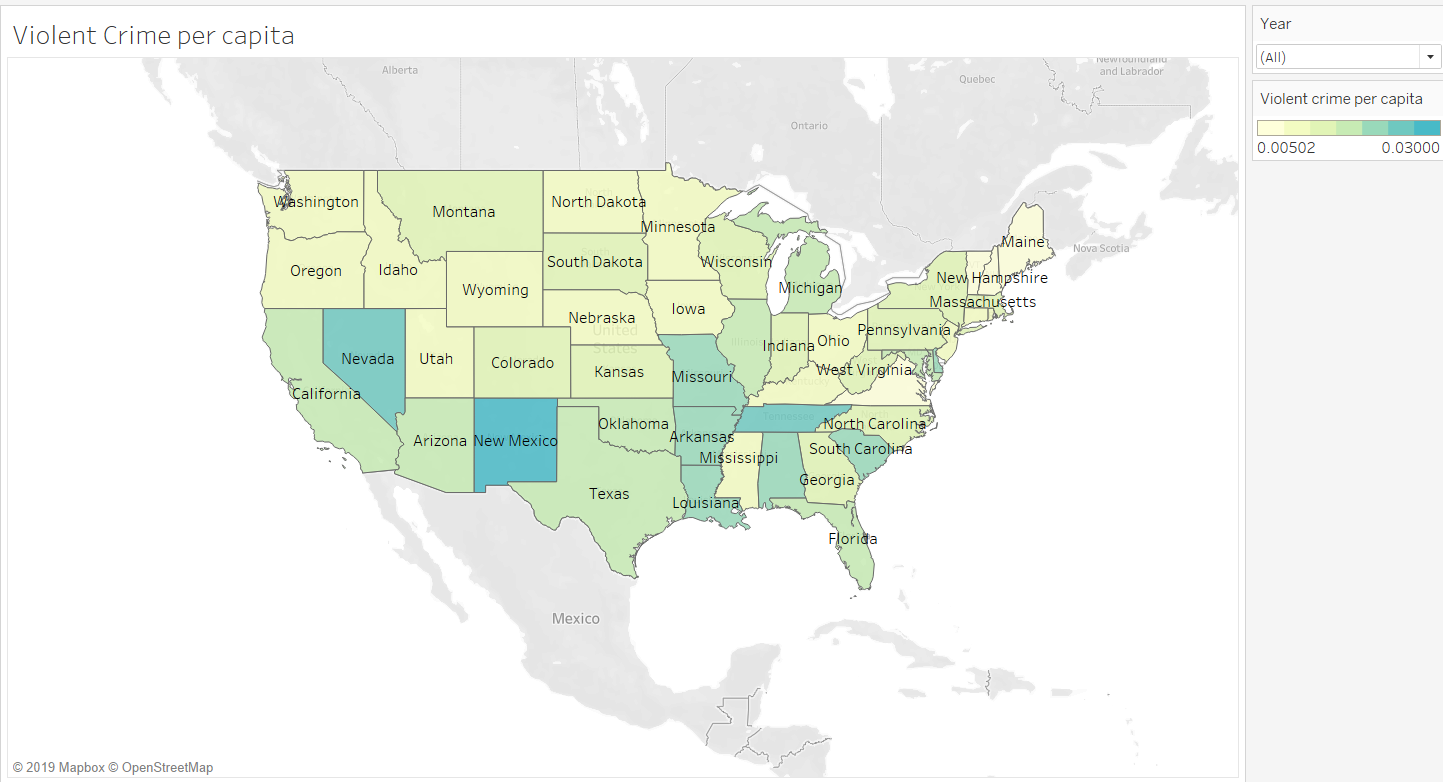
1. Unemployment Rate by state:

The map below shows the unemployment rates of the states in America for the years between 2014 and 2017. The darker shades of Teal denote higher rates of unemployment while the lighter shades show lower rates as indicated by the color legend on the side. The years of data can also be filtered using the year filter on the right. It can be observed that Nevada and New Mexico have higher rates of unemployment while North and South Dakota have lower rates of unemployment.



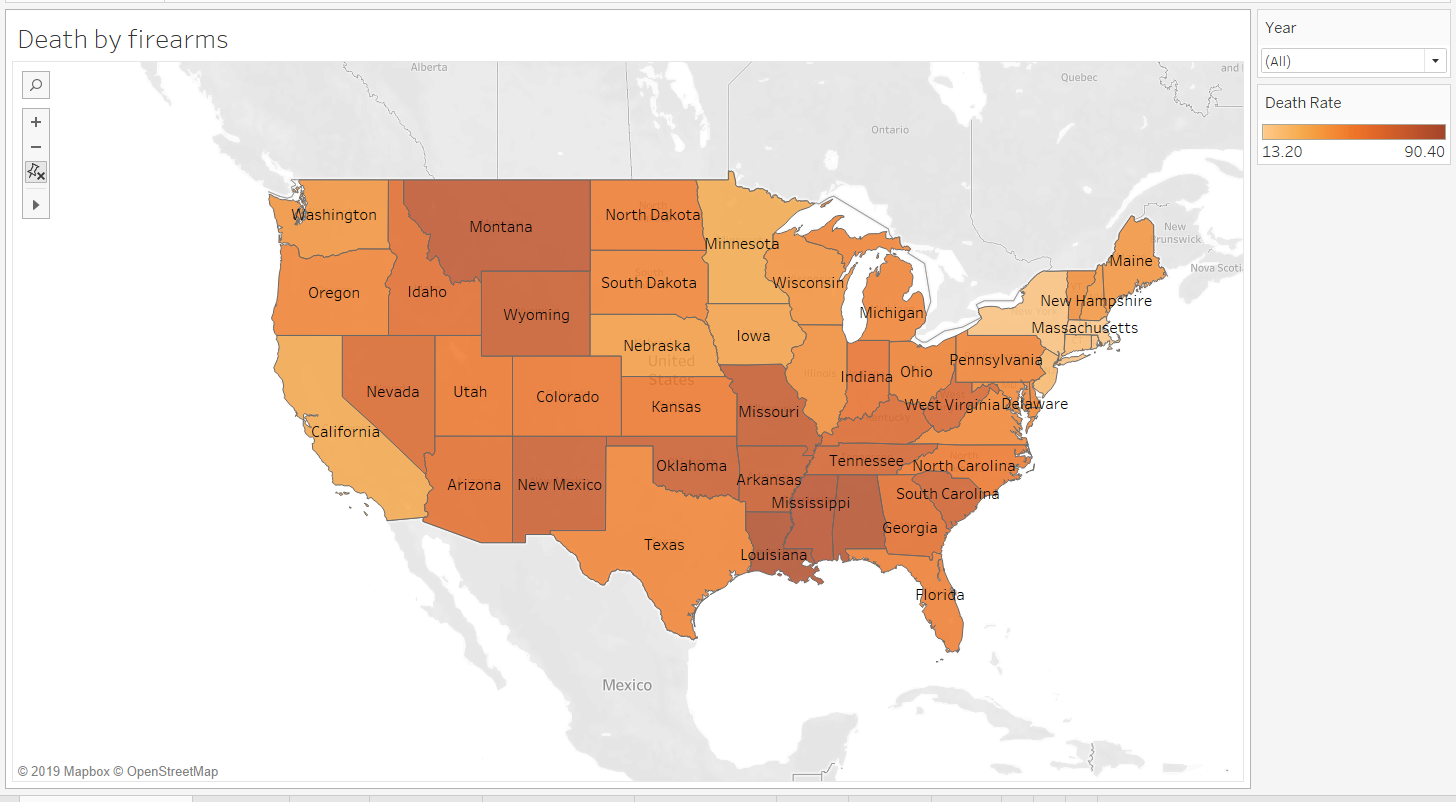
1. Violent crime per capita by state:

The map below shows the rate of violent crimes in the states in America for the years between 2014 and 2017. The darker shades of green denote higher rates of violent crimes while the lighter shades show lower rates as indicated by the color legend on the side. The years of data can also be filtered using the year filter on the right. It can be observed that states like Nevada, New Mexico, Missouri, Arkansas and Louisiana have higher rates of violent crime while states in the north have relatively lower rates. Other maps also follow the same design philosophy.



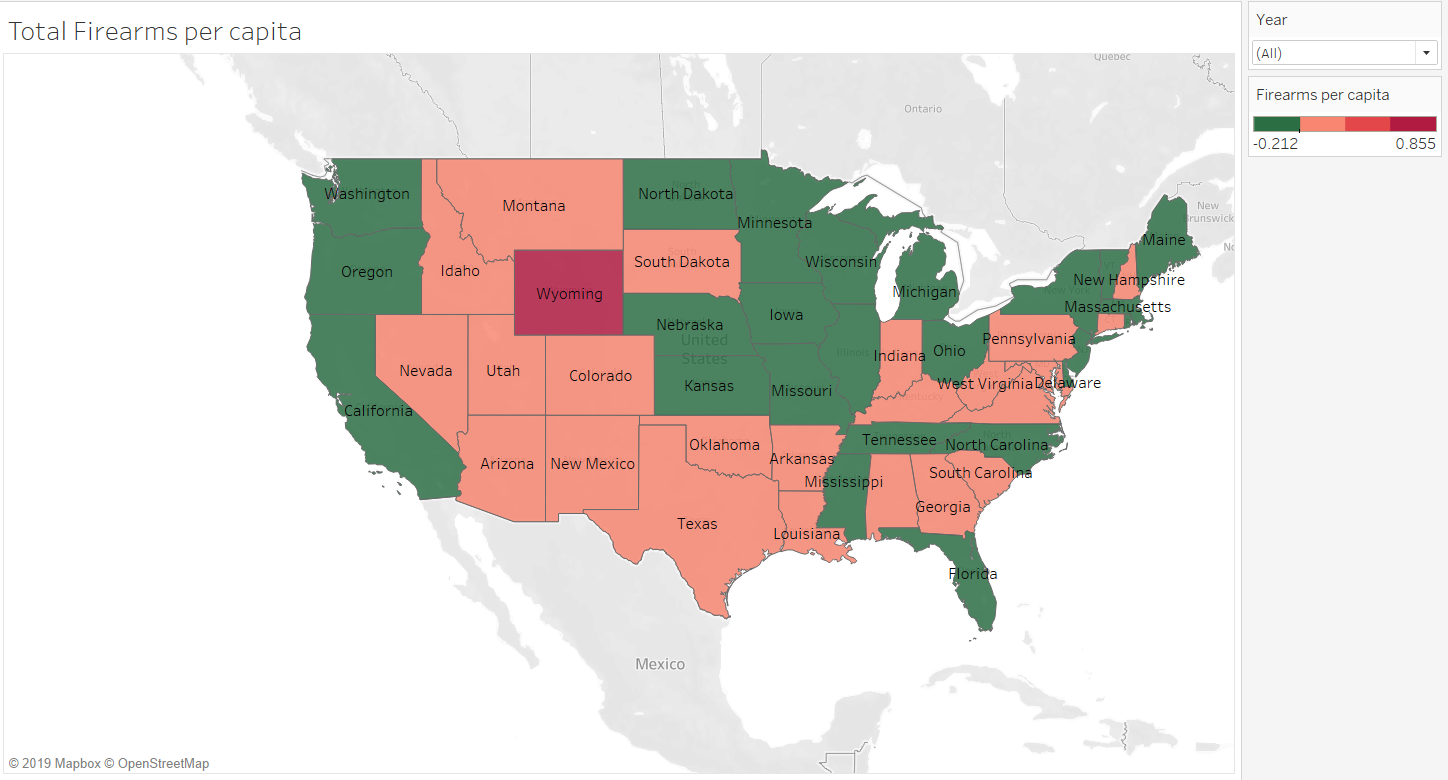
1. Firearm deaths by state:

It can be observed from the following map that states like Montana, Mississippi and Louisiana have the highest rates of firearm deaths while New York has one of the lowest rates.



1. Firearms by state:

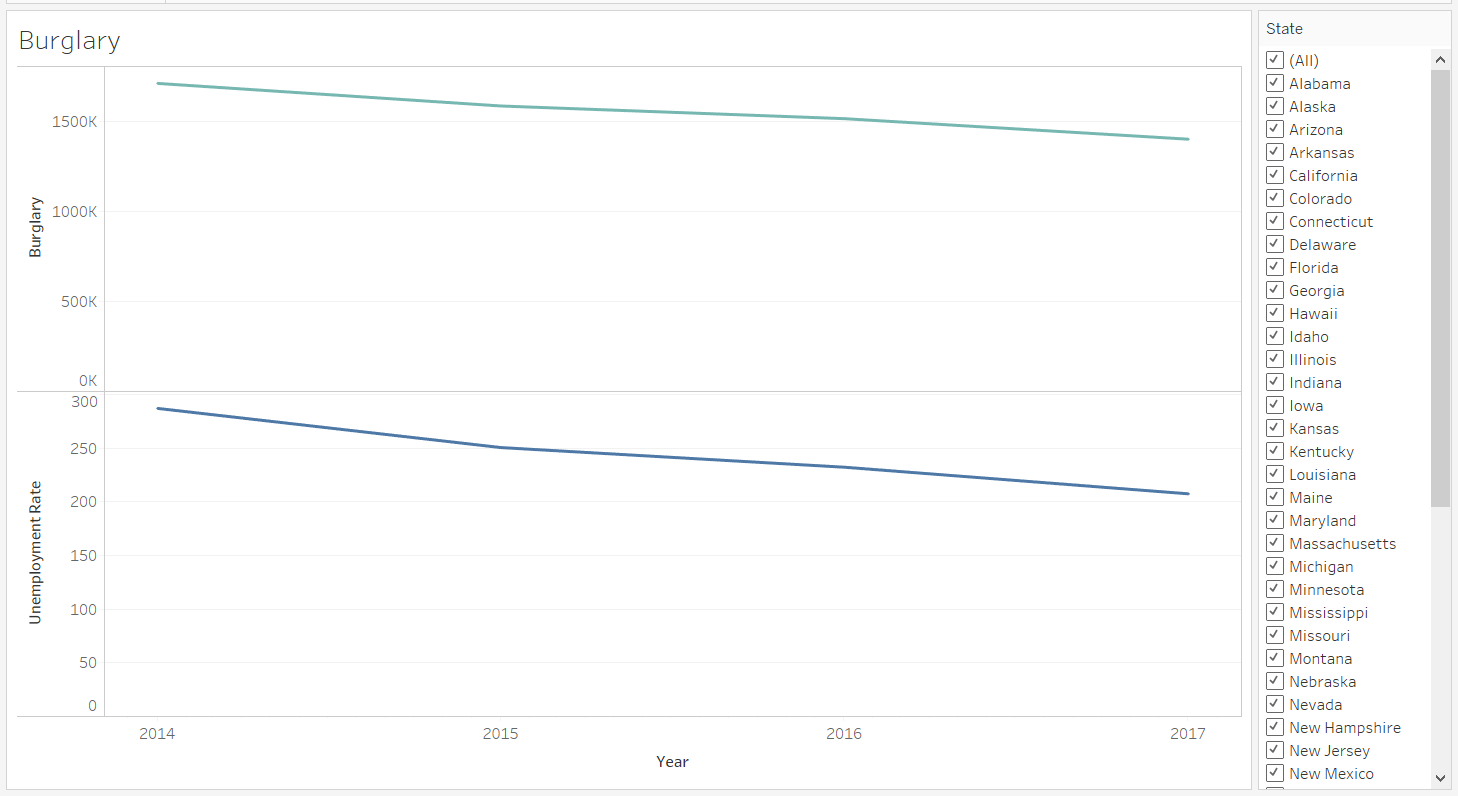
It can be observed from the following map that Wyoming has the highest number of firearms per capita while California and Florida are on the lower side.



Now, The Trend line comparisons below show the change in rates over the years between 2014 and 2017. The data shown in the charts below is the aggregate for all the states in America. This data can be filtered using the State filter card on the right.

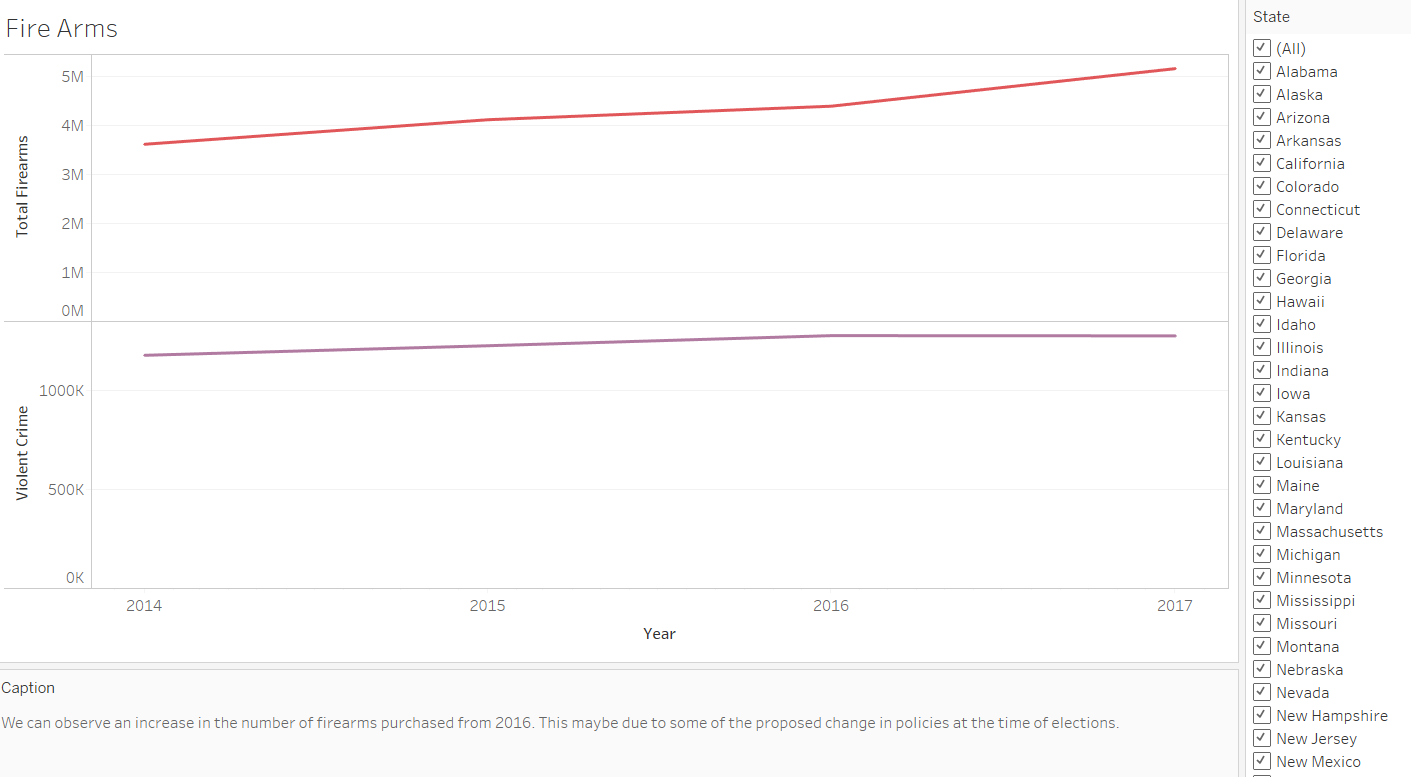
1. Burglary and Unemployment:

It can be observed that as the rate of unemployment has decreased over the years, the amount of burglary crimes has also declined.



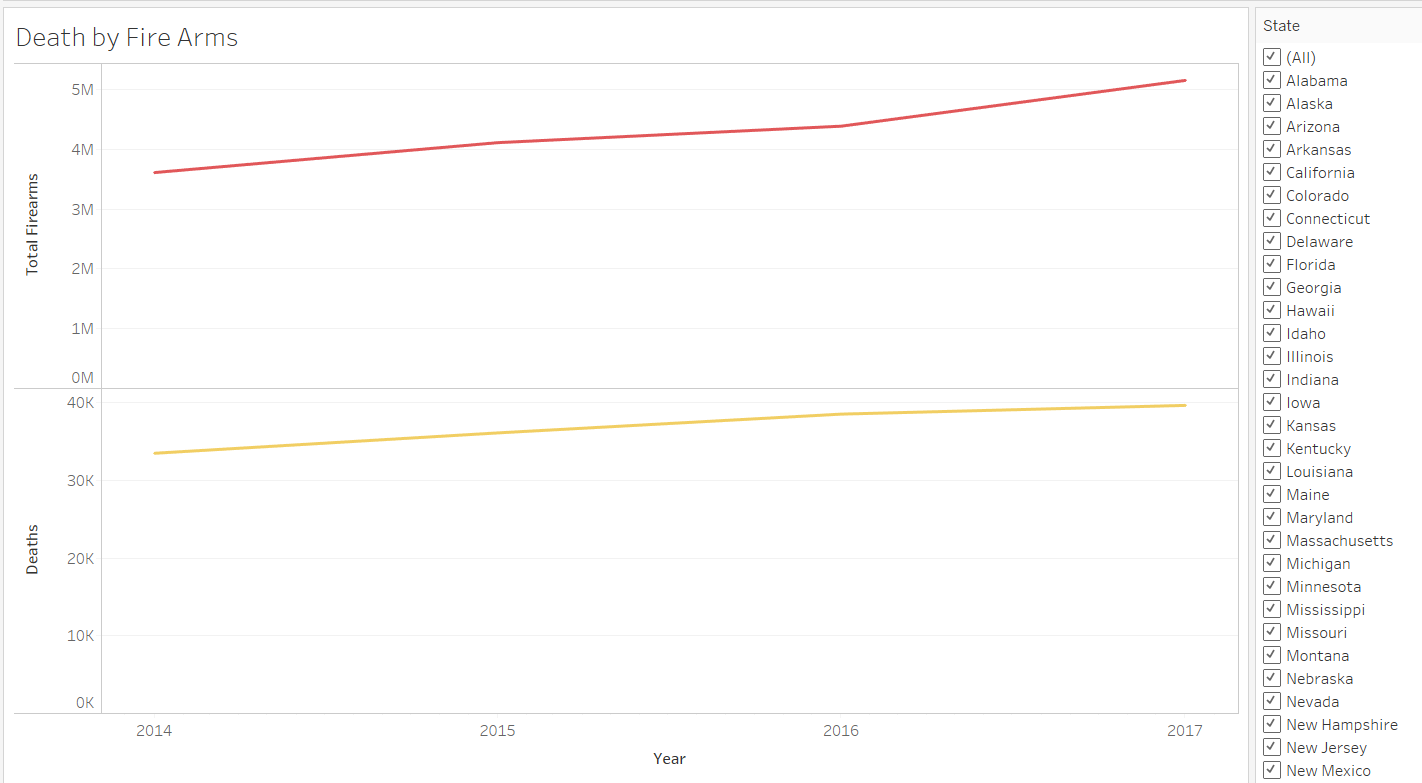
1. Firearms and Violent crimes:

There seems to not be a very high correlation between Firearms purchased and Violent crimes committed. However, that maybe due to high increase in firearm sales from 2016 due to some proposed political changes.



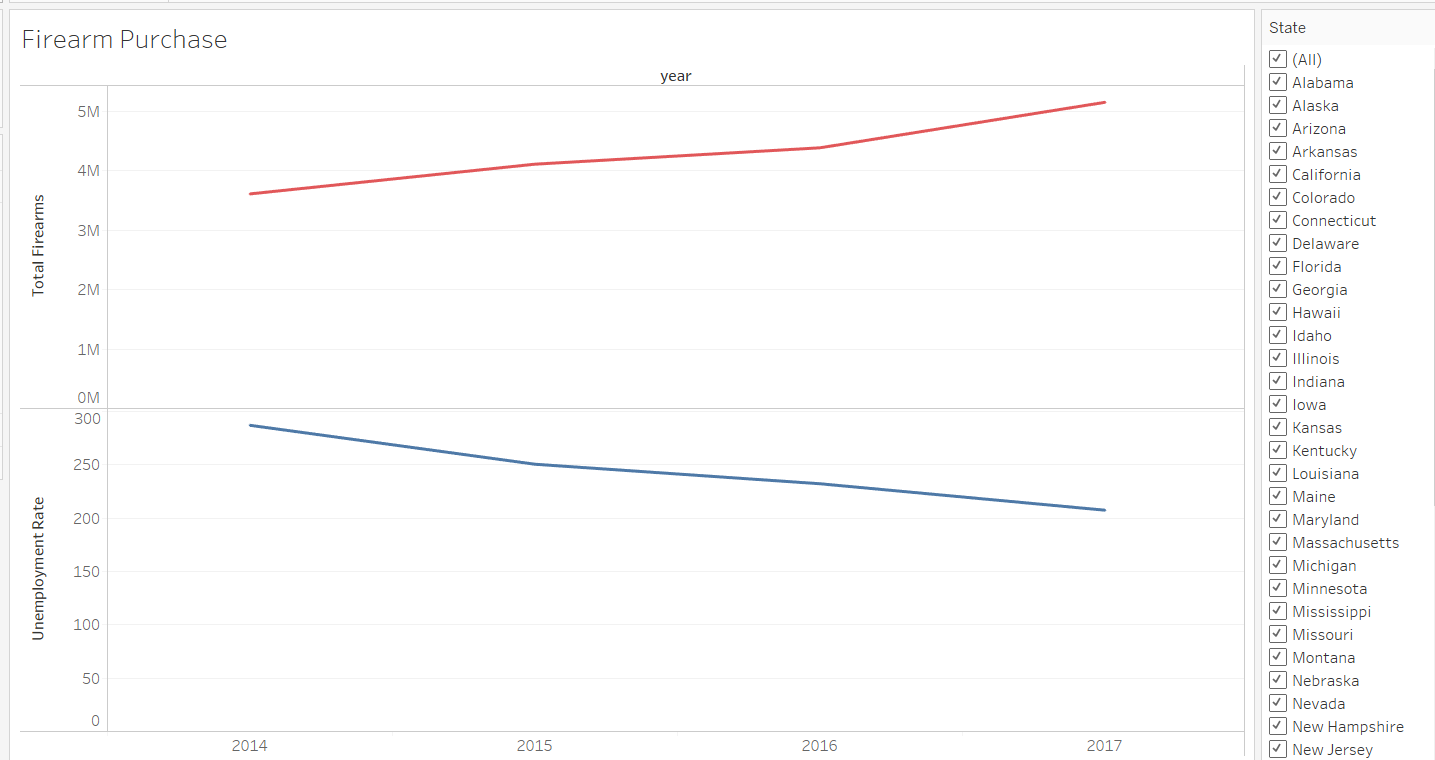
1. Firearms and death by firearms:

It can be observed that as the number of total firearms has increased over the years, the number of deaths by firearms has also increased.



1. Firearms and Unemployment:

It can be observed that as the rate of unemployment has decreased over the years, the purchase of firearms has increased.



Considerations for creating visualization:

* Only the visualizations that tell a story have been included. The analysis from these visuals can be used to determine the correlation and causation of crimes committed in the country and in which areas.
* These visualizations can be manipulated dynamically using filters to tailor to the needs of the end user.
* Visual considerations have also been kept in mind while creating these charts. Each Metric is represented by a specific color. It has been ensured that the colors selected are pleasing to the eye.
* These visualizations are scalable and can accommodate larger amounts of data to come in the future.

**Flowchart**

The figure below highlights the steps taken to complete the project. It starts with planning the project which includes brainstorming different topics, setting goals and deciding the timeline. It ends with interpreting the results based on the analysis.

A screenshot of a cell phone

Description automatically generated

**Instruction for the Code -** The code for the project has been written in an iterative manner as follows

1. First all the required packages are imported along with tabula which is a package in python that is used to fetch the data from a .pdf file into a dataframe.
2. A path has been saved as a variable where all the required dataset files have been stored. Hence, one must set a proper path before running the code into their own system. The path must have all the required dataset files. Datasets of Crime, Firearms, Unemployment and Mortality by firearms are to be downloaded and stored into the path from following sources:

* <https://crime-data-explorer.fr.cloud.gov/downloads-and-docs>
* [https://atf.gov/resource-center/data-statistics](https://www.atf.gov/resource-center/data-statistics)
* <https://www.bls.gov/lau/home.htm>
* <https://www.cdc.gov/nchs/pressroom/sosmap/firearm_mortality/firearm.htm>

1. New variables are created in which the data sets are being stored among which we include the path of the .pdf files from where the firearms data is to be fetched using tabula package.
2. Now, the code is written for the datasets in the order of ‘firearms’, ‘unemployment’, ‘crime’ and lastly ‘mortality by firearms’. It includes deleting the rows for null values, removing unwanted columns, re-casing and renaming the required columns in order to merge the data sets correctly and effectively. One may need not do any changes in-order to execute the code.
3. Lastly, we have included some lines of code to fetch analytical information from the final merged dataset. Such as correlation, covariance, summary, etc.

**Group Members:** Aqueel Jivan, Harsh Joshi, Khaza Towfik Akbar

**Topic:** Crimes and Firearms

**Datasets:** Crime, Firearms, Unemployment and Mortality by firearms

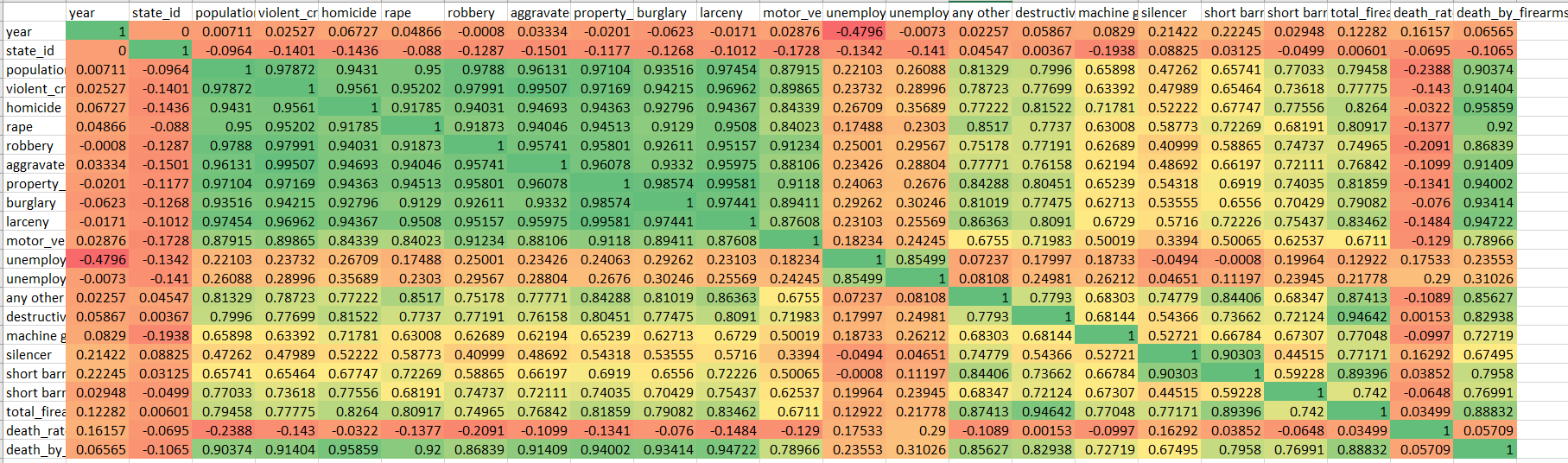
**Categorical Variables:** State, Year, State\_id, State\_abbr

**Continues Variables**: Unemployment\_rate, death\_rate

**Numeric Variables:**

* Types of crimes like Homicide, Larceny, rape, etc.
* Types of Firearms like Handgun, Machine guns, etc.
* Unemployment rate, rank
* Death by Firearms, rank

**Correlation Matrix**



(See Attachment with Submission for better clarity)

References:

1. <https://crime-data-explorer.fr.cloud.gov/downloads-and-docs>
2. [https://atf.gov/resource-center/data-statistics](https://www.atf.gov/resource-center/data-statistics)

3. <https://www.bls.gov/lau/home.htm>

4. <https://www.cdc.gov/nchs/pressroom/sosmap/firearm_mortality/firearm.htm>